ABSTRACT OF DISCLOSURE

A context of media content is represented by context description data having a hierarchical stratum. The context description data has the highest hierarchical layer, the lowest hierarchical layer, and other hierarchical layers. The highest hierarchical layer is formed from a single element representing content. The lowest hierarchical layer is formed from an element representing a segment of media content which corresponds to a change between scenes of video data or a change in audible tones. The remaining hierarchical layers are formed from an element representing a scene or a collection of scenes. A score corresponding to the context of a scene of interest is appended, as an attribute, to the element in each of the remaining hierarchical layers. A score relating to the time information about a corresponding media segment and a context is appended, as an attribute, to individual elements in the lowest hierarchical layer. In a selection step of a data processing method, the context of the media content is expressed, and one or a plurality of scenes of the media content is or are selected on the basis of the score of the context description data. Further, in the extraction step of the data processing method, only data pertaining to the scenes selected in the selection step are extracted.